Analysis of Classifiers for Performance Evaluation in Academic Sector using Classification Technique

Pooja Mishra [1], Neelam Sahu[2]

Dept. of IT, Dr. C. V. Raman University, Bilaspur (C.G.), India

---***---

Abstract:- An evaluation of teachers and students play very important role in academic sectors and institutions. Every academic sector like University, Colleges and schools are facing problem of good teachers and best results of students. The only way to improve the performance of academic sectors is evaluation of teachers and students. Data mining based classification technique is very important role for classify the students and teachers performance. In this research work we have used different classifiers for classifications of different level of teachers and student performance. QUEST gives better testing accuracy 68.42% for student performance. Similarly ANN gives 89.29% testing accuracy with Portuguese data set while QUEST gives 73.74% testing accuracy with Mathematics data set for Student Performance.

INTRODUCTION

In academic sector, evaluation of student and teachers performance is very challenging task for every institutes and organizations. The good teachers and good students makes healthy and good environment of any institutes and organization. The analysis and classification of good teachers and students play very important role to development of every organizations. In this research work there are various authors have worked in the field of teacher and student evaluation in academic sectors. Baradwaj B.K. et al., (2011) [7] have used decision tree method for classification and to predict the performance of the students. Kalpana J. et al., (2014) [6] have used data from a period of 2008-2013 and applied different clustering methods on the data and compared them as well. Ruby J. et al., (2016) [5] have used two datasets 1st is of 165 records long and the 2nd one is 396 records long. They have used MLP classification algorithm and got respectively 64.5% and 91.42% accuracy as an average of 5 runs. Gupta A. et al. (2015)[8] have suggested different classifiers for teaching assistant evaluation and achieved 41.05% of accuracy with J48 and Naive bayes classifier. J. Yang et al. (2011) [9] suggested SVM technique for Teaching Assistant Evaluation and compared the classifications techniques to develop various models for evaluation of teaching assistant evaluation.

1. METHODS

The techniques and tools are very important role in every field of research work. In this section we have explored different classifiers and data set used in this research work.

1.1 Decision Tree

Decision tree is one of the important data mining based classification technique that generates the rules to make the decision and classification. The main concept of decision tree [1] is to split data recursively into subsets so that each subset contains more or less homogeneous states of our target variable (predictable attribute). At each split in the tree, all input attributes are evaluated for their impact on the predictable attribute. When this recursive process is completed, a decision tree is formed. I have used CART, CHAID, QUEST and C4.5 decision tree technique that is used for classification and evaluation of student and teachers performance.

1.2 Artificial Neural Network

Artificial Neural networks [3] can be used for descriptive and predictive data mining. ANN is known as best classifier and is able to mine huge amount of data for classification. They were originally developed in the field of machine learning to try to imitate the neurophysiology of the human brain through the combination of simple computational elements (neurons) in a highly interconnected system. In this research work we have focus on learning rate and hidden layer. Learning rate updates the wheat at the time of learning and used to improve the performance of model. We have also used hidden layer one, two and three.

1.3 Bayesian Net

Bayesian classification [2] is based on Bayes' Theorem. Bayesian classifiers are the statistical classifiers. Bayesian classifiers can predict class membership probabilities such as the probability that a given tuple belongs to a particular class. Classification algorithms have found a simple Bayesian classifier known as naïve Bayes classifier to be comparable in performance with decision tree and selected neural network classifiers. Bayesian classifiers have also exhibited high accuracy and speed when applied to large dataset

1.4 Support Vector Machine (SVM)

Support vector machines (SVMs) [4] are supervised learning methods that generate input-output mapping functions from a set of labelled training data. The mapping function can be either a classification function (used to categorize the input data) or a regression function (used to estimation of the desired output). For classification, nonlinear kernel functions are often used to transform the input data to a high dimensional feature space in which the input data becomes more separable (i.e., linearly separable) compared to the original input space.

2. DATA SET

In this research work, we have used two different data sets: Teaching Assistant Evaluation(TAE) and Student Performance data set collected from UCI repository [10]. It contains 151 instances, 5 attributes and 1 class with three categories. Similarly student performance datasets divided into two distinct subjects as Portuguese and Mathematics. First dataset consist of 649 records and second dataset is consist of 395 record each of which is a multivariate dataset having 32 normal attributes and 1 special attribute that has a capability to become a label. Both datasets are modeled under binary classification problem.

3. RESULTS AND DISCUSSION

In this paper, we have used tow data two data set as student and teaching evaluation and analyzed and classify the performance of students and teachers using different data mining techniques. This experiment divided into two sections: (i) Teaching evaluation Performance (ii) Student Performance.

In first section ,TAE data set applied on the different classification techniques like CART, CHAID, QUEST, C5.0, ANN , SVM and Bayes Net to evaluate the performance of teachers . We have applied the 80% of data for training and 20% data for testing and compared the performance of classifiers where QUEST gives highest accuracy as 68.42% for teaching assistant evaluation.

Similarly we have also applied the same classifiers with same partition to evaluate the student performance. We have used two data set Portuguese and Mathematics and applied with different data mining based classification technique to evaluate the performance of teachers. In case of Portuguese data set ANN gives best accuracy 89.29% while QUEST gives 73.74% of accuracy in case of Mathematics data set.

Table 1 shows that accuracy of classifiers for TAE and Student Performance. Table 2: Confusion matrix of best model (QUEST) with TAE Data Set and Table 3 shows that Performance measures of best model (QUEST) with TAE Data Set. Table 4 shows that confusion matrix of best model with student performance Data Set. Table 5 shows that performance measures of best model with student performance data set.

Table 1: Accuracy of Classifiers for TAE and Student Performance

Classifier	TAE Data Set		Student Performance Data Set			
			Portuguese		Mathematics	
	Training	Testing	Training	Testing	Training	Testing
CART	69.03	44.74	88.02	86.43	78.72	71.72
CHAID	50.44	47.37	88.80	87.86	78.04	68.69
QUEST	58.41	**68.42**	85.85	87.14	71.96	**73.74**
C5.0	76.11	47.37	91.16	88.57	82.42	70.71

ANN	56.64	55.26	88.02	**89.29**	73.99	66.67
SVM	56.64	52.63	99.80	82.86	100	54.55
Bayes Net	69.91	57.89	93.91	80.0	88.85	58.59

Table 2: Confusion matrix of best model (QUEST) with TAE Data Set

Actual Vs. Predicted	Low	Medium	High
Low	7	5	0
Medium	3	10	0
High	1	3	9

Table 3: Performance measures of best model (QUEST) with TAE Data Set

Performance measures	Accuracy	Sensitivity	Specificity
Low	**87.88**	**58.34**	**84.61**
Medium	**92.10**	**76.92**	**68.00**
High	**89.47**	**69.23**	**60.00**

Table 4: Confusion matrix of best model with Student Performance Data Set

Actual Vs. Predicted	ANN with Portuguese data set		QUEST with Mathematics data set	
	Fail	**Pass**	**Fail**	**Pass**
Fail	7	12	11	18
Pass	3	118	8	62

Table 5: Performance measures of best model with Student Performance Data Set

ANN with Portuguese data set		QUEST with Mathematics data set	
Accuracy	**89.29**	Accuracy	**73.74**
Sensitivity	36.84	Sensitivity	37.93
Specificity	97.52	Specificity	88.57

4.   CONCLUSION AND FUTURE WORK

An evaluation of performance of teachers and students play very important role for every academicians and institutes. Every academics and institutes defines role of students and teachers based on their performance. Data mining based classification techniques play very important to analysis, classification and performance of classifiers. In this research work used various classifiers for analyzing and performance of students and teachers. The performance of clQUEST gives highest accuracy as 68.42% for teaching assistant evaluation. ANN gives best accuracy 89.29% with Portuguese data set while QUEST gives 73.74% of accuracy with Mathematics data set. In future we will develop the hybrid model to improve the performance of classifiers.

REFERENCE

[1]. A. K. Pujari, Data Mining Techniques, Universities Press (India) Private Limited, 4th ed., 2001.

[2]. H. Jiawei, K. Micheline, P. Jian, Data Mining Concepts and Techniques, Morgan Kaufmann, 2006.

[3]. P. Giudici, Applied Data Mining for Business and Industry, 2nd edition, John Wiley & Sons, 2009.

[4]. D. L. Olson and D. Delen, Advanced Data Mining Techniques, USA, Springer Publishing: ISBN: 978-3-540-76916-3,2008.

[5]. Ruby J. and David K., A study model on the impact of various indicators in the performance of students in higher education, IJRET International Journal of Research in Engineering and Technology, 3(5):750-755,2014.

[6]. J. Kalpana and K. Venkatalakshmi , Intellectual Performance Analysis of Students by Using Data Mining Techniques. International Journal of Innovative Research in Science, Engineering and Technology. 3(3):1922-1929,2014.

[7]. B. K. Baradwaj and S. Pal , Mining Educational Data to Analyze Students Performance. IJACSA, 2(6). 63-69, 2011.

[8]. A.Gupta, Classification Of Complex UCI Datasets Using Machine Learning And Evolutionary Algorithms", International journal of Scientific and technology research, 4: 94 , 2015.

[9]. J. Yang, H. Jiang and Honglei Zhang, Teaching Assistant Evaluation Based on Support Vector Machines with Parameters Optimization, Information Technology Journal, 10: 2140-2146,2011.

[10].   http://www.ics.uci.edu/~mlearn/databases/ (Browsing date: 18 Mar. 2018).